Electronic Colloquium on Computational Complexity, Report No. 153 (2005)

Testing Orientation Properties

Shirley Halevy* Oded Lachish† Ilan Newman‡ Dekel Tsur§

Abstract

We propose a new model for studying graph related problems that we call the *orientation model*. In this model, an undirected graph G is fixed, and the input is any possible edge orientation of G. A property is now a property of the directed graph that is obtained by a given orientation. The distance between two orientations is the number of edges that have to be redirected in order to move from one digraph to the other.

This model allows studying digraph properties such as not containing a forbidden (induced) subgraph, being strongly connected etc., for every underlying graph including sparse graphs. As it turns out, this model generalizes the standard, adjacency matrix model. That is, we show that for every graph property $\mathcal{P}$ of dense graphs there is a property of orientations that is testable if and only if $\mathcal{P}$ is. This model is also handy in some practical situations of networks, in which the underlying network is fixed while the direction of (weighted) links may vary.

We show that several orientations properties are testable in this model (for every underlying graph), while some are not.

*Department of Computer Science, Technion, Haifa 3200, Israel. Email: shirleyh@cs.technion.ac.il

†Department of Computer Science, University of Haifa, Haifa 31905, Israel. Email: loded@cs.haifa.ac.il Research is supported by CRI

‡Department of Computer Science, University of Haifa, Haifa 31905, Israel. Email: ilan@cs.haifa.ac.il Research supported in part by an Israel Science Foundation grant number 55/03.

§Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093

ISSN 1433-8092

1 Introduction

The goal of this paper is to introduce a new model for testing (di)graph related properties that we call the *orientation model*. We study this model and prove some relations to the standard (dense graph) model. In addition, we consider some specific properties of orientations in this model, for which we present efficient testing algorithms.

Property Testing for (di)graph properties has been first studies by [10], and since then attracted a considerable amount of attention (for a partial list of references and surveys see [1–5, 7, 9–16]). The standard and main model that was considered in the context of graph properties testing is the adjacency matrix model (also known as the "dense graph" model). In this model, a graph is represented by its adjacency matrix (namely, a Boolean vector specifying for every two vertices u, v whether there is an edge between u and v in the graph). The distance between two graphs in this model is measured by the number of entries in the matrix that should be altered in order to transform one graph to the other. That is, the distance between two graphs is the number of edges that need to be deleted/inserted in order to move from one graph to the other.

An (ϵ, q)-test for a (di)graph property $\mathcal{P}$ is a randomized algorithm that, on an unknown input graph G, probes q pairs of vertices from G. On each query, the tester receives the answer whether the pair is a (di)edge in G or not. It then should distinguish, with success probability of at least $2/3$, between the case that G has the property in question and the case that G is ϵ-far from having the property (meaning that at least ϵn^2 edges need to be deleted or inserted in order to transform G into a graph G' that has the property, where $n = |V(G)|$). A property is called *testable*, if there is an (ϵ, q)-test where $q = q(\epsilon)$ (that is, q does not depend on the graph size). This model is only suitable for the study of properties of dense graphs, as every ϵ-sparse graph (a graph with at most ϵn^2 edges) is ϵ-close to the empty graph (in the sense that at most ϵn^2 edges have to be deleted so that it becomes the empty graph). Property-testing for sparse graphs was first considered by Goldreich and Ron [12], who presented the *adjacency list* model for *bounded-degree* graphs. In this model, a graph is given by its *adjacency lists* and the distance between two bounded-degree graphs is measured by the number of entries in the adjacency lists that should be modified to transform one graph to the other. Other variants that allow to study graphs of varying edge densities were considered by [15] and [14].

We propose a new model for testing of digraph related properties that studies the properties of orientations. An orientation of the underlying graph $G = (V, E)$, is a digraph that is obtained by replacing each edge in E by a directed edge. In the orientations model one knows the undirected graph $G = (V, E)$, which we call the underlying graph, and gets an orientation of G as an input. A property is then a property of orientations, e.g. being strongly connected, being acyclic, not containing certain directed configurations etc. Orientations of a graph G are represented in the natural way, namely by a vector $D : E(G) \longrightarrow \{0, 1\}$ specifying the direction of each edge in G. The distance between two orientations is then measured by the Hamming distance between the two vectors representing them. Meaning that, in order to transform one digraph to the other, we allow to change the direction of edges but not to delete or insert edges. An (ϵ, q)-test for a property of G-orientations is defined analogously to the standard model. A query now is to an edge in the underlying graph to which the answer is the direction of the edge in the input orientation.

The motivation behind this new model is four fold: First, there are some real world application that directly call for this or similar models. Consider a given communication network, the underlying graph at a given time is fixed, while the direction (upload/download capacities) of information flow through communication lines are the varying/controlled parameters. In this setting, like many other cases of fixed topology transportation networks such as railway or roads networks, properties of interest are properties of the directed (sometimes weighted) digraph and how such properties might change when a small number of edges are redirected (reweighted).

The second motivation is that the orientation model allows the study of many properties of digraphs and orientations in an interesting framework that was not done before. It allows the study of properties of sparse graphs in a 'clean' setting which avoids the 'representation issues' (as in [12, 14]). As it turns out, the ability to test a given orientation property may rely heavily on the structure of the underlying graph (as exemplified by some of our results). Hence, the study of graph properties in the orientation model, may reveal interesting combinatorial structures of the underlying graph. In turn, the proofs might have their graph theoretic counter part that can be of interest outside the context of property testing.

A third motivation is that, although this model seems to be unrelated to the standard graph model, it turns out that it generalizes it for both digraphs and undirected graphs in the following sense; We show, via a simple reduction, that for every property $\mathcal{P}$ there is a corresponding property of orientations $\vec{\mathcal{P}}$ (and an underlying graph) such that $\mathcal{P}$ is testable in the dense graph model if and only if $\vec{\mathcal{P}}$ is testable in the orientation model.

This relation to the standard model, may open a new avenue of research: We know that properties that are defined by a finite collection of forbidden subgraphs are testable in the standard model. However, the dependency on the distance parameter ϵ is huge as a result of using the regularity lemma. One of the features of this model, as dealing with sparse graphs too, is that using the regularity lemma is not a natural choice. This may result in new algorithms (for 'old' graph properties) that are more efficient in term of the dependence in ϵ from those obtained by using the regularity lemma for the dense graphs model [1] (for 'forbidden subgraphs' properties , this would be a major breakthrough in the area of graph property testing).

Another feature that makes the model intriguing and is exemplified by our results, is that the natural algorithm that always work for dense graphs properties, does not necessarily work here: In the standard model, for any property $\mathcal{P}$ that is defined by a finite collection of forbidden subgraphs, it holds that if G is ϵ-far from $\mathcal{P}$, then it contains many instances of the forbidden subgraphs [1]. This was generalized by [13] to any graph property (for any fixed n).

The situation here is different; we show that, in the testable property of being *source-drain*-free, it could be the case that an orientation is ϵ-far from the property while having an insignificant amount of sources and drains. This is also true for the forbidden subgraph case: We show that there is a digraph that is ϵ-far from not having a directed triangle while having only $O(n)$ triangles. Thus, algorithms for this model are potentially more interesting than those for graph properties in the standard model. Finally, as we will show, the orientation model allows for interesting positive results, with sophisticated algorithms.

We note that there are other related variants that should be studied. For example, one can consider the problem of testing subgraphs that was suggested by M. Krivelevich. In this model we fix an underlying graph and the input is a subgraph (or alternatively, an induced subgraph) given by a 0/1 labeling of the edges (vertices) of the underlying graph. The distance between two inputs is just the hamming distance of the representing characteristic vectors. A property now is any collection of subgraphs (this obviously generalizes the standard model in which the fixed underlying graph is the complete graph). However, by the same reduction we present, orientation properties generalize this model as well (although the properties may become somewhat less natural). There are other iteresting extensions in these directions, e.g properties of coloring (of vertices or edges) of the underlying graph by more than 2 colors. This naturally gives rise to properties of colorings, partitions etc. We don't conduct here a study of such models, but do resort to such property in our lower bound in Section 6.

Results in this draft: Section 3 presents a reduction from the dense graph model to the orientation model, followed by some additional features of the orientation model (Section 4). Then, in Section 5, we study the orientation property of being *drain-source-free*. A directed graph D, is source-drain-free if there is no vertex $v \in V(D)$ whose edges are all either going out of v (which makes v a source) or all going into v (which makes v a drain). The property of being source-drain-free is quite natural in transportation networks. It also seems to be a first step towards testing strong connectivity. We show that this property is ϵ-testable for every underlying graph G. Namely, for every graph G there is a (1-sided error) test that on any input orientation $\vec{G}$ queries $q = exp(poly(1/\epsilon))$ edges (on which it receives as an answer the direction of the edge) and distinguishes, with success probability of 2/3, between the case that $\vec{G}$ is *source-drain-free* and the case that it is ϵ-far from being so. Note that, although the result holds for every graph, the undirected graph G is fixed in advanced and is not part of the input. We show that this property is testable both for the hamming distance and for the weighted hamming distance too (in the weighted hamming distance each edge has a weight and the distance between two orientations is the sum of weights over all the edges that are directed differently in both graphs).

In Sections 6, we study the property of being H-free for a fixed forbidden graph H. We prove that for underlying graphs that are of bounded degree, and for H that either does not have a source, or does not have a drain, this property is testable with $poly(1/\epsilon)$ queries. We also show that this is in some sense best possible in the following respect: We show that the property of being P_3-free[1] is highly non-testable even for bounded degree graphs. Thus, the positive result above cannot be generalized for every H even for the bounded degree case. We also note that if we relax the restriction on the underlying graphs to be of bounded average degree (rather then bounded maximum degree) then the property of being C_6-free[2] is not testable with $poly(1/\epsilon)$ queries.

Finally, in Section 7, we study the property of being *strongly connected*. We show positive results for a large class of underlying graphs.

2 Preliminaries

We use $\{u, v\}$ for undirected edges and (u, v) for edges directed from u to v. Let G be an undirected graph, and denote by n the number of vertices in $V(G)$. We say that a directed graph D is an *orientation* of the graph G, or in short a G-orientation, if we can derive G' from G by replacing each undirected edge $\{u, v\} \in E(G)$ by either the directed edge (u, v) or the directed edge (v, u) (but not both). We use the notation $\vec{G}$ to denote an orientation of G, and call G the *underlying graph* of $\vec{G}$. When we refer to more than one orientation we use indexing; for example, $\vec{G}_1$ and $\vec{G}_2$ are orientations of the graph G.

Note that the model is defined to include simple digraphs with no parallel or anti-parallel edges. This results in a cleaner model in terms of input representation. However, all the results that we present can be generalized to the model in which we allow parallel edges.

Given two G-orientations $\vec{G}_1, \vec{G}_2$, the relative distance between $\vec{G}_1$ and $\vec{G}_2$, denoted by $dist(\vec{G}_1, \vec{G}_2)$, is the number of edges in $E(G)$ such that their direction in $\vec{G}_1$ is different than their direction in $\vec{G}_2$. That is, $dist(\vec{G}_1, \vec{G}_2) = |\{\{u, v\} \in E(G) | (u, v) \in E(\vec{G}_1),\ (v, u) \in E(\vec{G}_2)\}|$. Given a G-orientation $\vec{G}$ and an edge $(u, v) \in E(\vec{G})$, the G-orientation $\vec{G_1}$ obtained from $\vec{G}$ by flipping the edge $(u, v) \in E(\vec{G})$, is defined by $E(\vec{G_1}) = (E(\vec{G}) \setminus \{(u, v)\}) \cup \{(v, u)\}$.

A property $\mathcal{P}_G$ of G-orientations is a subset of all the G-orientations. We say that an orientation $\vec{G}$ satisfies the property $\mathcal{P}_G$ if $\vec{G} \in \mathcal{P}_G$. The distance of $\vec{G}_1$ from the property $\mathcal{P}_G$ is defined by $dist(\vec{G}_1, \mathcal{P}_G) = \min_{\vec{G}_2 \in \mathcal{P}_G} dist(\vec{G}_1, \vec{G}_2)$. We say that $\vec{G}$ is ϵ-far from $\mathcal{P}_G$ if $dist(\vec{G}, \mathcal{P}_G) \geq \epsilon |E(G)|$, otherwise we say that $\vec{G}$ is ϵ-close to $\mathcal{P}_G$. In all notations we omit the subscript G when it is obvious

[1] P_3 is the directed path of length 3.
[2] C_6 is the 6 dicycle.

from the context. Next, we define the notion of an ϵ-test for a property of G-orientations.

Definition 2.1. *Let G be a fixed undirected graph and P be a property of G-orientations. An (ϵ, q)-test for $\mathcal{P}$ is a randomized algorithm that accesses the input G-orientation via edge queries. On query $e \in E(G)$ it receives the orientation of e in $\vec{G}$. The algorithm may ask at most q queries to $\vec{G}$ and should distinguish with success probability $2/3$ between the case that $\vec{G} \in P$ and the case that $\vec{G}$ is ϵ-far from P. If the algorithm never errs on an input $\vec{G} \in P$, we say that it has a one sided error.*

The complexity of an (ϵ, q)-test is the number of queries q. We say that a property is testable if it has an $(\epsilon, q(\epsilon))$-test, where $q(\epsilon)$ is independent of the graph size.

3 Reduction

In this section, we show that the orientation model that we present in this work in fact generalizes the dense graph model.

Theorem 3.1. *For every property $\mathcal{P}$ of (di)graphs on n vertices there is an undirected graph G and an orientation property $\vec{\mathcal{P}}$ such that $\mathcal{P}$ is (ϵ, q)-testable if and only if $\vec{\mathcal{P}}$ is (ϵ, q)-testable.*

We note that the undirected graph G in the theorem depends only on n and is the same for every $\mathcal{P}$. The number of vertices, as well as the number of edges in G, will be shown to be $\Theta(n^2)$.

Proof. Let $\mathcal{P}$ be a digraph property for graphs of n vertices which we want to test in the standard model. The fixed graph $G = G_n = (V_G, E_G)$ is defined as follows: $V_G = \{v_{i,j} | 1 \leq i \leq j \leq n\}$ and $E_G = \{\{v_{i,i}, v_{i,j}\}, \{v_{i,j}, v_{j,j}\}|$ for each $1 \leq i < j \leq n\}$. Clearly, $|V_G| = \frac{n \cdot (n+1)}{2}$ and $|E_G| = n \cdot (n-1)$. In particular, the average degree of a vertex in G is less than 4.

For every digraph on n vertices $D = ([n], E)$ we define the following orientation G_D of G: For every $1 \leq i < j \leq n$, if $(i, j) \in E(D)$, then $(v_{i,i}, v_{i,j}), (v_{i,j}, v_{j,j}) \in E(G_D)$. Similarly, if $(j, i) \in E(D)$, then $(v_{j,j}, v_{i,j}), (v_{i,j}, v_{i,i}) \in E(G_D)$. Otherwise, $(v_{i,i}, v_{i,j}), (v_{j,j}, v_{i,j}) \in E(G_D)$. Note that, there is a directed edge $(i, j) \in E(D)$ if and only if there is a path of length 2 from $v_{i,i}$ to $v_{j,j}$ in G_D. We now define the orientation property $\vec{\mathcal{P}}$ of G-orientations as follows: $\vec{\mathcal{P}} = \{G_D \mid D \in \mathcal{P}\}$.

We first prove that if $\vec{\mathcal{P}}$ is testable then so is $\mathcal{P}$. Indeed we claim that for every digraph on n vertices, $dist_O(G_D, \vec{\mathcal{P}}) = dist_A(D, \mathcal{P})$, where $dist_O$ denotes the distance in the orientation model and $dist_A$ denotes the distance in the adjacency-matrix model. To see this note that ,by the construction, $D \in \mathcal{P}$ implies that $G_D \in \vec{\mathcal{P}}$. Also, if one needs to delete/insert k edges to D so to enter $\mathcal{P}$, then there are k corresponding edges whose redirection will cause G_D to enter $\vec{\mathcal{P}}$. This proves that $dist_O(G_D, \vec{\mathcal{P}}) \leq dist_A(D, \mathcal{P})$. A similar reasoning proves the other direction.

Note that, the above proves that the absolute distance between the corresponding properties is maintained by the reduction. We note that this also proves that the relative distance is maintained up to a low order term. This is true as the relative distance in the adjacency matrix model is defined by $dist_A/n^2$, while for the orientation model it is $dist_O/|E_G| = dist_O/n^2 + o(1)$.

Finally, given an ϵ-test for $\vec{\mathcal{P}}$ in the orientation model, there exists an appropriate test for $\mathcal{P}$ in the standard model using the same query complexity. This is due to the fact that each query of an edge in G_D can be trivially simulated by a single query of the corresponding edge in D.

This completes the proof that if $\vec{\mathcal{P}}$ is testable then so is $\mathcal{P}$. The reverse implication is similar with the minor complication that a test for $\vec{\mathcal{P}}$ needs to consider also G-orientations that do not correspond to any digraph D on n vertices (and hence are not in $\vec{P}$). However, one could easily see that such orientations are either easy to test or are close to an orientation that corresponds to a digraph D on which all our previous reasoning work. $\square$

We note that the above reduction transforms the natural properties of being $\mathcal{H}$-free for a class of forbidden subgraphs $\mathcal{H}$ (finite or infinite) to an orientation property that is also defined by a forbidden collection of subgraphs. For example, the orientation property that corresponds to the property of being di-triangle-free is being C_6-free, where C_6 is the directed 6-cycle.

4 Some additional features of the orientation model

In [13] it is shown that if a graph property is testable by a 1-sided error algorithm then there is also a *non-adaptive* test that picks at random a subset of vertices, queries all edges between them and rejects only if the the subgraph obtained has a certain predefined property. In particular, it was already shown by [1] for properties that are defined by a finite set of forbidden (induced) subgraphs, that if a graph is ϵ-far from the property, then a constant fraction of its (induced) subgraphs (of appropriate size) are isomorphic to one of the forbidden subgraphs. Hence, a witness can be found just by sampling at random.

In the orientation model, we don't know if the property of being di-triangle free is testable (as opposed to the dense graph model [1,4]). However, we do know that if it is testable, the test cannot be the standard test — namely, the number of triangles in an ϵ-far orientation might be only $O(n)$ (compared to $\Omega(n^3)$ possible triangles) as the following example shows.

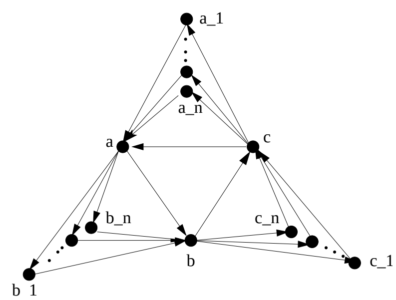

Figure 1: a graph that contains 1 di-triangle but is far from being triangle free.

Let $\vec{G}$ be the G-orientation that is shown in Fig 1. The only directed triangle in $\vec{G}$ is the triangle $T = (a, b, c)$. However, any change of a single edge in it, creates $\Omega(n)$ different triangles. As $|E(G)| = O(n)$ it is clear that $dist(\vec{G}, \text{triangle-free}) = \Omega(n)$. This example can be generalized to a dense graph G with $\Omega(n^2)$ edges and an orientation that has $O(n)$ di-triangles but is 1/9-far form being triangle-free.

5 Drain-Source-Freeness

Let $D = (V, E)$ be a digraph. We say that $v \in V$ is a source if there is no edge $(x, v) \in E$ (that is, there exists no edge in D that is directed towards v). Similarly, $v \in V$ is a drain if there is no edge $(v, x) \in E$. Define $Drains(D)$ to be the set of all drains in D, and $Sources(D)$ to be the set of all sources in D.

Definition 5.1. *A directed graph D is called drain-source-free, if $Drains(D) \cup Sources(D) = \emptyset$. Similarly, D is drain-free (respectively, source-free), if $Drains(D) = \emptyset$ (respectively, $Sources(D) = \emptyset$). We say that D is 2DSF if there is no source or drain of degree 2 in D.*

In what follows, we show that the orientation property of being *drain-source-free* (as well as *drain-free* and *source-free*) is testable for any underlying graph. We present here a test for the case in which the input G-orientations are assumed to be $2DSF$ (this is the main technical part). The general test, which is a simple modification of the test here, is presented in Appendix C.

An important feature of those tests is that they work for weighted graphs as well. In the weighted case, the distance between two G-orientations $\vec{G}_1$, $\vec{G}_2$ is defined to be $\sum_{e \in E(\vec{G}_1) \triangle E(\vec{G}_2)} weight(e)$, where $E(\vec{G}_1) \triangle E(\vec{G}_2)$ is the set of all edges that have different orientation in the two orientations.

Remark 5.1.: The tests we describe are for the case that the underlying graph is connected, has at least one vertex of degree greater than 2 and does not contain any vertices of degree 1. Extending the tests to deal with general underlying graphs is trivial.

We first present an ϵ-test that distinguishes between orientations that are drain-source-free and those that are 2DSF (that is, have no drain or source of degree 2), and are ϵ-far from being drain-source-free. The test is non adaptive and has a 1-sided error, that is, it rejects a G-orientation only if it finds a drain/source in it.

The test only looks for drains/sources of degree smaller than a specified polynomial in $\frac{1}{\epsilon}$. We later prove that the impact of sources and drains of higher degree on the distance is negligible. For $v \in V$, denote by $deg_G(v)$ the degree of the vertex v. We formally state the bound on the degree.

Definition 5.2. *Let G be an undirected graph. We say that $u \in V(G)$ is* ϵ-relevant *if* $3 \leq deg_G(u) \leq 36/\epsilon$ *and* ϵ-heavy *if $deg_G(u) > 36/\epsilon$. Denote by $Heavy_G$ the set of all the ϵ-heavy vertices in G.*

For every two vertices $u, v \in V(G)$, denote by $dist_G(u, v)$ the length of the shortest undirected path from u to v in G. For every vertex $v \in V(G)$, and for every integer $r \leq n$, define $B_G(v, r)$ to be the subgraph of G spanned by the set of all vertices u such that $dist(u, v) \leq r$. Denote by $\tilde{B}_G(v, r)$ the subgraph of G obtained from $B_G(v, r)$ by removing every edge $\{x, y\}$ such that $dist_G(x, v) = dist_G(y, v) = r$.

Definition 5.3. *For every graph $G = (V, E)$, vertex $v \in V$, and $\epsilon > 0$, define $rad_{G,\epsilon}(v)$ to be the minimal integer r such that at least one of the following requirements is satisfied:*
- *$\tilde{B}_G(v, r)$ is not a tree.*
- *$\tilde{B}_G(v, r)$ contains heavy vertices.*
- *there are at least $36/\epsilon$ vertices at distance exactly r from v.*

Claim 5.4. *For every graph $G = (V, E)$ and every vertex $v \in V$, $\tilde{B}_G(v, rad(v))$ contains at most $2(36/\epsilon)^2$ relevant vertices. In particular, a path from v to a vertex x in $\tilde{B}_G(v, rad(v))$ contains at most $2(36/\epsilon)^2$ relevant vertices.*

Proof. Let h be the number of vertices at distance $rad(v) - 1$ from v. By Definition 5.3, $h < 36/\epsilon$. Consequently, since $\tilde{B}_G(v, rad(v) - 1)$ does not contain heavy vertices and the degree of every relevant vertex is at most $36/\epsilon$, there are at most $h \cdot (36/\epsilon)$ vertices at distance $rad(v)$ from v. Since $\tilde{B}_G(v, rad(v) - 1)$ is a tree and the degree of every relevant vertex is at least 3, $\tilde{B}_G(v, r)$ contains at most $2(36/\epsilon)^2$ relevant vertices. □

The test, assuming that the input orientations are 2DSF, is as follows:

Algorithm 5.1.
Input: $\epsilon, \vec{G}$.
Repeat $16/\epsilon$ times:

1. *Randomly choose an edge $\{u, v\} \in E(G)$ such that $v \notin Heavy$.*
2. *For every relevant vertex z in the graph such that $dist(v, z) < rad(z)$ query all incident edges to z and reject if z is a drain/source.*

If no drain or source was found, return PASS.

Theorem 5.5. *Algorithm 5.1 is an $(\epsilon, (1/\epsilon)^{O((\frac{1}{\epsilon})^2)})$-test for Drain-Source-Free.*

Proof. We first determine the query complexity of the algorithm. We start by bounding from above the number of relevant vertices z than can be selected in one run of step 2 of algorithm 5.1. Since $dist(v,z) < rad(z)$, by Claim 5.4, the shortest path between v and z contains at most $2(36/\epsilon)^2$ relevant vertices. Hence, there are less than $(36/\epsilon)^{2(36/\epsilon)^2}$ such paths and consequently there are at most $(36/\epsilon)^{2(36/\epsilon)^2}$ such relevant vertices. Since step 1 is repeated $16/\epsilon$ times and we query all edges only for such relevant vertices, the total number of query is bounded by $(\epsilon, (1/\epsilon)^{O((\frac{1}{\epsilon})^2)})$.

Clearly, the test does not reject a drain-source-free orientation. It remains to prove that every orientation $\vec{G}$ that is 2DSF and is ϵ-far from being Drain-Source-Free is rejected with high probability.

By Claim 5.6, which we state and prove further on, if $\vec{G}$ is ϵ-far from Drain-Source-Free then it is either $\epsilon/3$-far from Drain-Free or it is $\epsilon/3$-far from Source-Free. Without loss of generality, assume that $\vec{G}$ is $\epsilon/3$-far from Drain-Free. Then, by Lemma 5.7 below, the probability that for a vertex v that is selected in step 1 of algorithm 5.1 there exists a vertex $z \in (Drains(\vec{G}) \cup Sources(\vec{G})) \setminus Heavy$ such that $v \in \tilde{B}_G(z, rad(z))$ is at least $2\epsilon/9$. The proof follows. $\square$

The following claim states that if a G-orientation is far from being Drain-Source-Free then it is either far from being Source-Free or far from being Drain-Free. Note that a similar relation between the sum of distances form two peoperties and the distance to their union is not necessarily true in general. The proof of the claim appears in Appendix A.

Claim 5.6. *If $\vec{G}$ is ϵ-far from Drain-Source-Free then either $\vec{G}$ is $\epsilon/3$-far from Drain-Free or $\vec{G}$ is $\epsilon/3$-far from Source-Free.*

The following is the main technical lemma used in the correctness proof of Algorithm 5.1.

Lemma 5.7. *Let $\vec{G} \in$ 2DSF be ϵ-far from Drain-Free then, for at least a $2\epsilon/3$ fraction of the edges $\{u,v\} \in E(G)$ such that v is ϵ-relevant, there exists a vertex $x \in Drains(\vec{G}) \setminus Heavy$ such that $v \in \tilde{B}_G(x, rad(x))$.*

The proof of the above lemma is presented in Appendix B and relies on the following lemma.

Lemma 5.8. *For every graph G and every G-orientation $\vec{G} \in$ 2DSF, $dist(\vec{G}, \text{Drain-Free}) \leq \frac{4\cdot|E(G)|}{\Delta_{Dr}(\vec{G})}$, where $\Delta_{Dr}(\vec{G}) = \min_{x \in Drains(\vec{G})} deg_G(x)$ (that is, $\Delta_{Dr}(\vec{G})$ the minimal degree of a drain in $\vec{G}$).*

6 Testing of being H-free

In this section we further study the general property of being H-free where H is a finite fixed digraph. As will be shown in Theorem 6.3, it is not true that the property of being H-free is testable for every underlying graph and for every H, hence we don't have here a positive result as for drain-source freeness. However, we present (essentially best possible) results about testing H-freeness in bounded degree underlying graphs.

Theorem 6.1. *Let H be a fixed digraph that is either source-free or drain-free. Then, for any $O(1)$-bounded degree graph G, H-freeness of G-orientations is ϵ-testable by $poly(1/\epsilon)$ queries.*

Proof. The test is the analogue of the 'canonical tester' for the case of orientations which is sampling a node at random, sampling a large enough ball around this node and rejecting based on the subgraph obtained. We first observe following (proof can be found in Appendix E).

Observation 6.2. *Let $G = (V, E)$ be an undirected graph such that $deg(v) < \Delta$ for each $v \in V$ and let H be a source-free digraph. If $dist(\vec{G}, H\text{-free}) \geq \epsilon n$ then at least $\epsilon n/\Delta$ of the vertices in G appears in a copy of H in $\vec{G}$.*

Observation 6.2 immediately implies that the following is a 1-sided error ϵ-test for being H-free: Repeat the following $10\Delta/\epsilon$ times. Choose a random vertex $v \in V(G)$ and query all the edges that are adjacent to a vertex in the ball $B(r, v)$ for $r = |V(H)|$. Reject if a copy of H is found.

Clearly, the query complexity is $O(1/\epsilon)$ and the test has 1-sided error. Assume that $\vec{G}$ is ϵ-far from being H-free and H is source free. By Observation 6.2, the probability that a random chosen v is in a copy of H in $\vec{G}$ is at least ϵ/Δ. In addition, if v is in an H-copy in $\vec{G}$, then this copy will be discovered by the test. Thus, the claim follows. □

We note that all the above applies also to the property of having no *induced* subgraph that is isomorphic to H.

Theorem 6.1 presents positive results only for a restricted class of forbidden subgraph H and bounded degree underlying graphs. One may wonder what happens if we remove one of those restriction. Let P_3 be a directed path of length 3 (and 4 vertices). We show, in Theorem 6.3, that testing of being P_3 free for a certain family of underlying graphs (of bounded degree) needs linear number of queries. Note, that since P_3 has a source and a drain, this implies that Theorem 6.1 is best possible in that respect.

As a side remark, we note that testing for being P_2 free is actually very easy: it is meaningful only for underlying graphs that are bipartite (as otherwise, any orientation contain a P_2) and it is done by just sampling $O(1/\epsilon)$ edges.

Theorem 6.3. *For an infinite sequence of $n \in \mathbb{N}$ there is an underlying graph G on n vertices and degree bounded by 10, for which any test for the orientation property of being P_3 free requires $\Omega(n)$ queries.*

Proof. Let G be a graph and let (A, B, C) be a partition of the vertex set into three (possibly empty) parts, we say that the partition defines a *proper* 3-coloring if there is no edge in G such that both its end-points are in the same part. Similarly, a coloring is proper if there is no edge such that both its end-points are colored by the same color. We denote by i-path a directed path of length i. We first note that if a G-orientation is P_3-free then G is 3-colorable. To see that consider a P_3-free orientation, $\vec{G}$. We may assume that $\vec{G}$ has no directed triangles (as if it does, each di-triangle is disconnected from the rest of the graph). Let $V_0 = \{v \in V(G) \mid \textit{there is a } 2-\textit{path starting at } v \textit{ in } \vec{G}\ \}$. Similarly, let $V_2 = \{v \in V(G) \mid \textit{there is a } 2-\textit{path ending at } v \textit{ in } \vec{G}\ \}$. Let $V_1 = \{v \in V(G) \mid v \textit{ is a middle vertex in } 2-\textit{path in } \vec{G}\}$. Note that since $\vec{G}$ is triangle free and has no path of length 3 it follows that $\{V_0, V_1, V_2\}$ are pair-wise disjoint and each being an independent set in G. Also note that every vertex $u \in V(G) - \cup_{i=0}^{2} V_i$ is either a source, or a drain. Putting each source vertex in $V(G) - \cup_{i=0}^{2} V_i$ into either V_1 or V_0 (and if put into V_1, putting also drains adjacent to it, that are in $V(G) - \cup_{i=0}^{2} V_i$ into V_2) and putting each remaining drain into V_1 or V_2, results in a partition that defines a proper 3-coloring of G. We call any such coloring a coloring consistent with the orientation $\vec{G}$.

The above discussion proves that for any P_3-free orientation any consistent coloring defines a proper 3 coloring of G. The other direction is also true; any 3-coloring of G defines a P_3-free orientation by orienting all edges from lower color class to higher color class.

For a 3-colorable undirected graph G let $coloring(G)$ be the property containing all proper 3-coloring of G. The results in [8] implicitely[3] implies that there is an ϵ (and an infinite sequence of $n \in \mathbb{N}$) for which there is a 3-colorable graph G on n vertices such that any ϵ-test for the property $coloring(G)$ requires $\Omega(n)$ queries. It can also be seen that the graphs obtained from the proof of [8] can be made to be of degree bounded by 10.

We now show that $coloring(G)$ is reducible to the orientation property of being P_3-free. Indeed let G be a 3-colorable graph and $c : V(G) \longrightarrow \{0, 1, 2\}$ be any 3 coloring of $V(G)$ (not necessarily a proper 3-coloring). We define the following G-orientation $\vec{G}_c$: For each edges going in between two color classes direct the edge from the lower to the higher color. For each edge inside a color class we choose the orientation arbitrary. To prove that this is a reduction between two testing problems we should prove that the following three requirements hold:

1. If $c : V(G) \longrightarrow \{0, 1, 2\}$ is a proper coloring of G then $\vec{G}_c$ is a P_3-free orientation.

2. If $c : V(G) \longrightarrow \{0, 1, 2\}$ is ϵ-far from being a proper 3-coloring then $\vec{G}_c$ is ϵ'-far from being P_3-free from some ϵ' that should not depend on n.

3. If there is a test for being P_3-free then there is a test for being a proper coloring using roughly the same amount of queries.

Indeed the first item follows immediately the discussion above, since the mapping between the inputs is such that $\vec{G}_c$ is P_3-free if and only if c is a proper coloring. This also immediately implies the 3rd item as the mapping is local, thus an alleged test for P_3-free can be simulated by a test for coloring by just querying the colors of the two endpoints of an edge which will determine the orientation of the edge.

The 2nd item is not true in general. However, as we only aim for a lower bound on testing orientations, it is enough to have that the 2nd item is correct on ϵ-far colorings that are 'hard-to-test'. To be specific, we show that if c is ϵ-far from being a proper coloring and moreover, it has only $o(n)$ monochromatic edges, then $\vec{G}_c$ is ϵ-far from being P_3-free. This is enough as coloring that have $\Omega(n)$ monochromatic edges are easy to test just by sampling an edge at random. Thus, the hard to test coloring must be those for which the 2nd item holds and thus the lower bound follows. The proof of the 2nd item for the hard to test coloring can be found in Appendix F. □

We now note that for testing H-freeness for H that is either drain-free or source-free, the assumption of being bounded degree cannot be simply relaxed to the assumption that G has $O(n)$ edges (equivalently, a bounded average degree).

Observation 6.4. *There is a graph with $O(1)$-average degree for which there is no ϵ-test for the property of not containing a directed 6-cycle, that has $poly(1/\epsilon)$ query complexity.*

Proof. Alon et al. [4] showed that in the standard (dense graph) model, a test for the property of not containing a (directed) triangle cannot be of $poly(1/\epsilon)$ query complexity (and independent of n). By Theorem 3.1, for every directed graph $D = (V, E)$, there exists an undirected graph $G_D = (V', E')$ with constant average degree, such that any ϵ-test for being di-6-cycle-free for $\vec{G_D}$ induces an ϵ-test in the standard model of the same complexity for D being triangle free. □

[3] [8] shows that there is a linear size $3CNF$ that is highly not testable. The standard reduction from $3CNF$ to coloring implies the above.

7 Testing being Strongly Connected

In this section we present partial results on testing the orientation property of being *strongly connected*. The basic tool that we use is inspired by [6]. We comment that we still do not know if this property is testable for every underlying undirected graph.

We start with a few observations. Obviously, we may assume that the underlying graph $G = (V, E)$ is 2 edge-connected as otherwise no orientation is strongly connected. We can also assume that $|E| = O(n)$ as otherwise $\vec{G}$ is ϵ-close to be strongly connected for any G-orientation $\vec{G}$ (by simply orienting a minimal 2-edge connected subgraph of G to be strongly connected). Thus it is enough to consider 2-edge connected sparse graphs.

Let D be a directed graph. We refer to the DAG (directed acyclic graph) $ST(D)$, that is defined on the strongly connected components of D in the standard way. A *source component* of D is a strongly connected component C in D that corresponds to a source vertex in $ST(D)$ (in other words, every edge between a vertex in C and a vertex in $V(D) \setminus C$ is directed away from C). A *drain component* of D is defined similarly.

The following is an observation used in [6] and which we will recurrently use.

Observation 7.1. *Let $\vec{G}$ be a G-orientation. If $\vec{G}$ has at least $\Omega(n)$ sources or drains components then a source or drain component of $\vec{G}$ can be found in $O(1)$ queries.* □

We don't know if the orientation property of being "strongly connected" is testable in general. However, we prove that for a rather large family of graphs that is defined below, the property is testable.

Definition 7.2. *An undirected 2-edge connected graph $G = (V, E)$ is called efficiently-Steiner-connected if for every $\delta > 1/\log n$ and $S \subseteq V$ with $|S| \leq \delta^2 n$ there is a connected subgraph $T = (V', E')$ such that $S \subseteq V'$ and $|E'| \leq 10\delta n$*

Note that we may always assume that T is a Steiner tree spanning S. We also note that for our purpose, the constant 10 and the function $\log n$ in the definition are quite arbitrary (and could be replaced by other constant and any $f = w(1)$).

Theorem 7.3. *If G is efficiently-Steiner-connected then the G-orientations property of being strongly connected is testable. (We omit the proof for lack of space).*

We note that any 'slightly expanding' graph is strongly-Steiner-connected, while for example, the cycle is not. A simple application of Theorem 7.3 is given by the following two theorems.

Theorem 7.4. *Let G be the $\sqrt{n} \times \sqrt{n}$ two dimensional grid. Then the property of being strongly connected is testable for G-orientations. (The proof is given in the Appendix 7.4).*

Theorem 7.5. *Let G be an expander graph then the G-orientations property of being strongly connected is testable. (We omit the proof for lack of space).*

References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs, FOCS* 1999, pp. 656–666.

[2] N. Alon and A. Shapira *A Characterization of Easily Testable Induced Subgraphs, SODA* 2004, pp. 935–944.

[3] N. Alon and A. Shapira *Every Monotone Graph Property is Testable, STOC* 2005, pp. 128–137.

[4] N. Alon and A. Shapira *Testing Subgraphs in Directed Graphs, STOC* 2003, pp. 700–709.

[5] A. Bogdanov, K. Obata, and L. Trevisan, *A lower bound for testing 3-colorability in bounded-degree graphs, FOCS,* 2002, pp. 93-102.

[6] M. A. Bender and D. Ron, *Testing properties of directed graphs: acyclicity and connectivity,* Random Struct. Algorithms 20(2):184–205 2002

[7] E. Fischer, *The art of uninformed decisions: A primer to property testing, The Computational Complexity Column of The bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[8] Eli Ben-Sasson and Prahladh Harsha and Sofya Raskhodnikova, *Some 3CNF properties are hard to test, STOC* 2003, pages 345–354.

[9] O. Goldreich, *Combinatorical property testing – a survey, In: Randomized Methods in Algorithms Design*, AMS-DIMACS pages 45–61, 1998 .

[10] O. Goldreich, S. Goldwasser, and D. Ron, *Property testing and its connection to learning and approximation, Journal of the ACM*, 45(4):653–750, 1998 (a preliminary version appeared in FOCS 1996).

[11] O. Goldreich and D. Ron, *On testing expansion in bounded-degree graphs. In Electronic Colloquium on Computational Complexity* 7(20), 2000.

[12] O. Goldreich and D. Ron, *Property testing in bounded degree graphs, STOC* 1997, pp. 406–415.

[13] O. Goldreich and L. Trevisan, *Three theorems regarding testing graph properties, FOCS* 2001, pp. 302–317.

[14] T. Kaufman, M. Krivelevich, and D. Ron, *Tight bounds for testing bipartiteness in general graphs, RANDOM-APPROX* 2003, pp. 341–353.

[15] M. Parnas, and D. Ron, *Testing the diameter of graphs, RANDOM-APPROX* 1999, pp. 85–96.

[16] D. Ron, *Property testing (a tutorial), In: Handbook of Randomized Computing* (S.Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolin eds), Kluwer Press (2001).

Appendix

A Proof of Claim 5.6

Assume that $\vec{G}$ is $\epsilon/3$-close to Drain-Free and $\epsilon/3$-close to Source-Free. We will show that this immediately implies that $dist(\vec{G}, \text{Drain-Source-Free}) < \epsilon|E(G)|$.

We claim that for every $\vec{H}$ and $\vec{H}' \in$ Drain-Free such that $dist(\vec{H}, \vec{H}') = dist(\vec{H}, \text{Drain-Free})$ we have that $Sources(\vec{H}') \subseteq Sources(\vec{H})$. We prove this claim after we show how it leads to the required contradiction.

According to this claim if we select $\vec{G}' \in$ Drain-Free such that $dist(\vec{G}, \vec{G}') = dist(\vec{G}, \text{Drain-Free}) < \epsilon \cdot |E(G)|/3$ then $Sources(\vec{G}') \subseteq Sources(\vec{G})$. Again according to the claim if we select $\vec{G}'' \in$ Source-Free such that $dist(\vec{G}', \vec{G}'') = dist(\vec{G}', \text{Source-Free})$ then $Drains(\vec{G}'') \subseteq Drains(\vec{G}')$. Since $Sources(\vec{G}') = \emptyset$ we get $\vec{G}'' \in$ Drain-Source-Free. By the triangle inequality we get $dist(\vec{G}, \text{Drain-Source-Free}) \leq dist(\vec{G}, \vec{G}'') \leq 2 \cdot dist(\vec{G}, \text{Drain-Free}) + dist(\vec{G}, \text{Source-Free}) < \epsilon \cdot |E(G)|$.

Let $\vec{H}, \vec{H}'$ be G-orientations such that $\vec{H}'$ is drain-free and such that $dist(\vec{H}, \text{Drain-Free}) = dist(\vec{H}, \vec{H}')$. Assume for contradiction that $Sources(\vec{H}')$ is not a subset of $Sources(\vec{H})$. Then there exists $v \in Sources(\vec{H}')$ such that $v \notin Sources(\vec{H})$. Hence there exists $u \in V(G)$ such that $(v, u) \in E(\vec{H}')$ and $(u, v) \in E(\vec{H})$. Let $\vec{H}''$ be the orientation we get from $\vec{H}'$ by flipping the edge (v, u). Obviously $dist(\vec{H}, \vec{H}'') < dist(\vec{H}, \vec{H}')$ and therefore it is enough to prove that $\vec{H}' \in$ Drain-Free in order to get a contradiction. Observe that all the edges that are adjacent to vertices in $V(G) \setminus \{u, v\}$ are directed in the same way in $\vec{H}'$ and $\vec{H}''$; hence, the vertices in $V(G) \setminus \{u, v\}$ are not in $Drains(\vec{H}'')$. The vertex u is not in $Drains(\vec{H}'')$ since $(u, v) \in E(\vec{H}'')$. The vertex v is not in $Drains(\vec{H}'')$ since it has degree at least 2 and was a source in $\vec{H}''$. $\square$

B Proof of Lemma 5.7 and Lemma 5.8

We first introduce some preliminary definitions and claims, then, in Subsection B.1, we prove Lemma 5.8, and in Subsection B.2 we conclude the proof of Lemma 5.7.

Given a drain v in $\vec{G}$, we will be interested in correcting v (by directing a path in the graph from it to a different vertex in the graph) without creating any new drain. Hence, we define the class of vertices and paths in the graph that are potential candidates for such a correction, and prove some features of the definitions that will be later used in the proof.

Definition B.1. *Given a G-orientation $\vec{G}$, we say that a vertex $v \in V(G)$ is a* target *vertex in $\vec{G}$, if v has out-degree of at least 2 in $\vec{G}$.*

Definition B.2. *Let $\vec{G}$ be a G-orientation. We associate with each vertex $v \in Drains(\vec{G})$ an arbitrary shortest path in G that starts in a target vertex x (with respect to $\vec{G}$) and ends in v. We denote this path by $Correction\text{-}Path_{\vec{G}}(v)$ and set $Correction\text{-}Dist_{\vec{G}}(v) = dist_G(v, x)$.*

Note: In definition B.2 the path is the shortest path in G (not in $\vec{G}$), though the definition is done with respect to a specific G-orientation $\vec{G}$.

Claim B.3. *For every $v \in Drains(\vec{G})$, the following holds: (1) The graph $B(v, r)$ is a directed tree for every $r < Correction\text{-}Dist_{\vec{G}}(v)$. (2) All the edges in $\tilde{B}(v, r)$ are directed towards v for every $r \leq Correction\text{-}Dist_{\vec{G}}(v)$.*

Proof. By induction on r. For $r = 1$ the claim is trivial.

Induction step: Assume that $\tilde{B}_G(v, r-1)$ is a directed tree whose edges are directed towards v.

(1) Assume that $B_G(v,r)$ for $r <$ Correction-Dist$_{\vec{G}}(v)$ is not a tree. This can only happen if one of the following is true. There exist $x, y \in V(G)$ such that $dist(x,v) = dist(y,v) = r$ and $\{x,y\} \in E(G)$ or there exist u, w, z such that $dist(w,v) = dist(u,v) = r-1$ and $dist(z,v) = r$ and $\{u,z\}, \{w,z\} \in E(G)$. It is immediate that one of the vertices x, u, w, z is a target in contradiction to the assumption that $r <$ Correction-Dist$_{\vec{G}}(v)$.

(2)Assume that not all the edges in $\tilde{B}(v,r)$ are directed towards v for $r \leq$ Correction-Dist$_{\vec{G}}(v)$. Then, there exists an edge $(y,x) \in E(\vec{G})$ such that $dist(x,v) = r$ and $dist(y,v) = r-1$. It is immediate that y is a target in contradiction to the assumption that $r \leq$ Correction-Dist$_{\vec{G}}(v)$. $\square$

Claim B.4. *For every two different drains u, v in $\vec{G}$, $\tilde{B}_G(v, \textit{Correction-Dist}_{\vec{G}}(v))$ and $\tilde{B}_G(v, \textit{Correction-Dist}_{\vec{G}}(u))$ are edge disjoint.*

Proof. To prove the above claim, it is enough to prove that $dist_G(u,v) \geq$ Correction-Dist$_{\vec{G}}(v)$ + Correction-Dist$_{\vec{G}}(u)$, for every $u, v \in Drains(\vec{G})$.

Let u, v be two drains in $\vec{G}$ and let T be a shortest path in G that connects u and v. Since both u and v are drains in $\vec{G}$, at least one of the vertices x along T has an out degree of at least 2 in $\vec{G}$. Hence, x is a target vertex in $\vec{G}$. Since $dist_G(v,x) \geq$ Correction-Dist$_{\vec{G}}(v)$ and $dist_G(u,x) \geq$ Correction-Dist$_{\vec{G}}(u)$, then Correction-Dist$_{\vec{G}}(v)$ + Correction-Dist$_{\vec{G}}(u) \leq dist_G(x,v) + dist_G(u,x) = dist_G(v,u)$. $\square$

However, it is easy to see that a correction of every drain in $\vec{G}$ to the closest target vertex (by directing the path from the drain to the target vertex) is not always possible without creating a new drain. A problem may rise when trying to correct more than one drain to the same target vertex. This situation is formally defined as follows.

Definition B.5. *A target vertex v in $\vec{G}$ is said to be* bad*, if for every edge $(v,x) \in E(\vec{G})$, there exists a drain u in $\vec{G}$ such that the correction path Correction-Path$_{\vec{G}}(u)$ ends in v and contains the edge (v,x). Define $Bad(\vec{G})$ to be the set of all bad vertices in $\vec{G}$.*

Definition B.6. *For every vertex $v \in V(G)$ and for every G-orientation $\vec{G}$ we define the* bad environment *of v, denoted by $Bad\_env_{\vec{G}}(v)$ as follows.*

- *If v is not a bad vertex in $\vec{G}$ then $Bad\_env_{\vec{G}}(v) = \emptyset$.*
- *Otherwise, set $Bad\_env_{\vec{G}}(v)$ to be the union of Correction-Path$_{\vec{G}}(u)$ over all drains u in $\vec{G}$ such that Correction-Path$_{\vec{G}}(u)$ contains v.*

In the proof we apply the following operation on *Bad_env* subgraphs.

Contracting a subgraph: The contraction transformation is given a graph G, a G-orientation $\vec{G}$, and a subset S of $V(G)$, and returns an undirected graph H and a H-orientation $\vec{H}$ that are obtained from G and $\vec{G}$ by replacing the vertices in S with a single vertex v_S. The output graphs H and $\vec{H}$ are formally defined as follows:

- $V(H) = (V(G) \setminus S) \cup \{v_S\}$.
- $E(H) = \{\{x,y\} \in E(G) \mid x, y \in V(G) \setminus S\} \cup \{\{x, v_S\} \mid x \in V(G) \setminus S,\ \exists y \in S \ s.t.\ \{x,y\} \in E(G)\}$.

- $E(\vec{H}) = \{(x,y) \in E(\vec{G}) \mid x,y \in V(G)\backslash S\} \cup \{(x, v_S) \mid x \in V(G)\backslash S,\ \exists y \in S \ s.t.\ \{x,y\} \in E(G)\}$.

Note that edges that are disjoint from S retain their direction in $\vec{H}$ as in $\vec{G}$, while for edges $\{x,y\}$ such that $x \notin S$ and $y \in S$, we impose the direction of (x, v_S) regardless of the direction of $\{x,y\}$ in $\vec{G}$. However, the following claim asserts that when applied to *Bad_env* the direction of the edges remains consistent.

Observation B.7. *For every bad vertex v in $\vec{G}$, and for every edge $\{x,y\} \in E(G)$ such that $x \in Bad\_env_{\vec{G}}(v)$ and $y \notin Bad\_env_{\vec{G}}(v)$, the edge $(y,x) \in E(\vec{G})$.*

Proof. Immediate from the definition of a correction path as the shortest path to a target vertex. Otherwise, x would have been a target vertex, and the correction path that contains x would have terminated in x and not in v. □

In our proof we contract all the *Bad_env*'s in a graph. To make sure that this operation is well defined we need the following simple observation.

Observation B.8. *For every two different bad vertices v and u in $\vec{G}$, $Bad\_env_{\vec{G}}(v) \cap Bad\_env_{\vec{G}}(u) = \emptyset$.*

B.1 Proof of Lemma 5.8

Let G and $\vec{G}$ be as in the lemma. We first construct from $\vec{G}$ a family of graphs that will be used in the proof that indeed $dist(\vec{G}, \text{Drain-Free}) \leq 4 \cdot |E(G)|/\Delta_{Dr}(\vec{G})$, where $\Delta_{Dr}(\vec{G})$ is the minimal degree of a drain in $\vec{G}$. This is constructed by the procedure 'correct' below.

Correct
Input: $\vec{G}$.
Set $\vec{H}_1 = \vec{G}$ and $i = 1$.
Repeat the following step while $Bad(\vec{H}_i) \neq \emptyset$.

- For each $v \in Bad(\vec{H}_i)$ contract the subgraph $Bad\_env_{\vec{H}_i}(v)$.
- Set H_{i+1} to the resulting graph and increase i by 1.

Set $k = i - 1$.

Note that by observation B.8 the contractions are well defined. We show that the above family of graphs satisfy certain conditions.

Claim B.9. *The following holds for every iteration of Procedure Correct.*

1. $|Drains(\vec{H}_{i+1})| \leq \left\lfloor \frac{|Drains(\vec{H}_i)|}{2} \right\rfloor$.
2. $Drains(\vec{H}_{i+1}) = Bad(\vec{H}_i)$.
3. $Bad(\vec{H}_i) \subset V(\vec{G})$.
4. $\Delta_{Dr}(\vec{H}_{i+1}) \geq (4/3)^i \cdot \Delta_{Dr}(\vec{G})$
5. $\sum_i \sum_{v \in Drains(\vec{H}_i)} Correction\text{-}Dist_{\vec{H}_i}(v) \leq 4|E(G)|/\Delta_{Dr}(\vec{G})$.

Proof. 1 and 2 are immediate from Definition B.6.

3. Assume towards a contradiction that for some i there exists $x \in Bad(\vec{H}_i)$ such that $x \notin V(\vec{G})$. By the definition of a bad vertex we have $x \notin Drains(\vec{H}_i)$. However, by the definition of procedure correct, x is the result of a contraction and hence by observation B.7 it is a drain of $\vec{H}_i$.

4. By the definition of *Bad_env*, $\Delta_{Dr}(\vec{H}_{i+1}) \geq 2 \cdot \Delta_{Dr}(\vec{H}_i) - 2$. Since $\vec{G}$ is free of degree 2 drains and $\Delta_{Dr}(\vec{H}_j) \geq 3$, then $\Delta_{Dr}(\vec{H_{i+1}}) \geq \frac{4}{3}\Delta_{Dr}(\vec{H}_i)$. Condition 4 immediately follows.

5. By claim B.4, for every two different drains u, v in $\vec{H}_j$, the two subgraphs $\tilde{B}_{\vec{H}_j}(v, \text{Correction-Dist}_{\vec{H}_j}(v))$ and $\tilde{B}_{\vec{H}_j}(v, \text{Correction-Dist}_{\vec{H}_j}(u))$ are edge disjoint. Therefore,

$$|E(G)| \geq |E(H)| \geq \sum_{v \in Drains(\vec{H}_j)} \left| E\left(\tilde{B}_{\vec{H}_j}(v, \text{Correction-Dist}_{\vec{H}_j}(v))\right)\right|. \tag{1}$$

In addition, by Claim B.3, $\tilde{B}_{\vec{H}_j}(v, \text{Correction-Dist}_{\vec{H}_j}(v) - 1)$ is a tree, and hence

$$\left| E\left(\tilde{B}_{\vec{H}_j}(v, \text{Correction-Dist}_{\vec{H}_j}(v))\right)\right| \geq deg_H(v) \cdot \text{Correction-Dist}_{\vec{H}_j}(v) \geq \Delta_{Dr}(\vec{H}_j) \cdot \text{Correction-Dist}_{\vec{H}_j}(v). \tag{2}$$

Equations 1 and 2 now imply

$$\sum_{v \in Drains(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v) \leq |E(G)|/\Delta_{Dr}(\vec{H}_j). \tag{3}$$

Summing up equation 3 over j and combining with condition 4 implies 5. ☐

Lemma B.10. *For every $i \leq k+1$,*

$$dist(\vec{H}_i, \textit{Drain-Free}) \leq \sum_{j=i}^{k+1} \sum_{v \in Drains(\vec{H}_j)} \textit{Correction-Dist}_{\vec{H}_j}(v).$$

Before proving the above lemma, we show why it completes the proof of Lemma 5.8. By Claim B.9 part 5, for every $i \leq k+1$, it holds that

$$\sum_j \sum_{v \in Drains(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v) \leq 4|E(G)|/\Delta_{Dr}(\vec{G}).$$

By Lemma B.10, $dist(\vec{H}_1, \text{Drain-Free}) \leq \sum_{j=1}^{k+1} \sum_{v \in Drains(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v)$. The desired claim now follows.

Proof of Lemma B.10. The proof is done by downwards induction.

For $i = k+1$ this is immediate because $Bad(\vec{H}_{i+1}) = \emptyset$ and hence the orientation we obtain by flipping all the edges in $\text{Correction-Dist}_{\vec{H}_{k+1}}(v)$ for every $v \in Drains(\vec{H}_{k+1})$ is Drain-Free. Therefore, $dist(\vec{H}_{k+1}, \text{Drain-Free}) \leq \sum_{j=k+1}^{k+1} \sum_{v \in Drains(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v)$.

Assume that the claim holds for $i+1$. That is, there exists a Drain-Free orientation $\vec{H}'_{i+1}$ such that $dist(\vec{H}'_{i+1}, \vec{H}_{i+1}) \leq \sum_{j=i+1}^{k+1} \sum_{v \in Drains(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v)$.

Define H_i to be the underlying graph of $\vec{H}_i$. We define a new H_i-orientation $\vec{H}'_i$ as follows. For every $\{x,y\} \in E(H_i)$ such that $x, y \in V(H_i) \setminus V(H_{i+1})$, if $(x,y) \in E(\vec{H}_i)$ then $(y,x) \in E(\vec{H}'_i)$ otherwise $(x,y) \in E(\vec{H}'_i)$ (one can view this as the result of flipping all the edges of Correction-Path$_{\vec{H}_j}(v)$ for every $v \in \mathit{Drains}(\vec{H}_i)$). For every $\{x,y\} \in E(H_{i+1})$ such that $x, y \in V(H_i) \cap V(H_{i+1})$ if $(x,y) \in E(\vec{H}'_{i+1})$ then $(x,y) \in E(\vec{H}'_i)$ otherwise $(y,x) \in E(\vec{H}'_i)$ (all edges that are "common" to $\vec{H}'_i$ and $\vec{H}_{i+1}$ are directed in the same way). For every $\{x,y\} \in E(H_i)$ such that $y \in V(H_i) \cap V(H_{i+1})$ and $x \in V(H_i) \setminus V(H_{i+1})$ and hence $x \in S = \mathit{Bad\_env}_{\vec{H}_i}(v)$ for some $v \in \mathit{Bad}(\vec{H}_i)$, if $(v_S, y) \in E(\vec{H}'_{i+1})$ then $(x,y) \in E(\vec{H}'_i)$, otherwise $(y,x) \in E(\vec{H}'_i)$ (each edge connecting a vertex from a bad environment to a vertex not in a bad environment is directed in the same way as the edge "corresponding" to it in $\vec{H}'_{i+1}$).

Observe that $\mathit{Drains}(\vec{H}') = \mathit{Bad}(\vec{H}_i)$. Consider the set $S = \mathit{Bad\_env}_{\vec{H}_i}(v)$ for some $v \in \mathit{Bad}(\vec{H}_i)$. By the definition of $\vec{H}'_i$, v is a drain in $\vec{H}'_i$. Since $\vec{H}'_{i+1}$ is drain-free, v_S is not a drain in $\vec{H}'_{i+1}$ and hence by the definition of $\vec{H}'_i$ there exists a vertex y in $V(S)$ that is a target in $\vec{H}'_i$. By the definition of S, there exists a (di)path in $\vec{H}'_i$ from y to x that goes only through vertices in S. It is easy to verify that flipping all the edges in such a (di)path for every $v \in \mathit{Drains}(\vec{H}'_i)$ results in a drain-free orientation $\vec{H}''_i$. Note that all flips where made in Correction-Path$_{\vec{H}_i}(v)$ for some $v \in \mathit{Drains}(\vec{H}_i)$ and hence

$$\mathit{dist}(\vec{H}_i, \vec{H}''_i) \leq \sum_{v \in \mathit{Drains}(\vec{H}_i)} \text{Correction-Dist}_{\vec{H}_i}(v) + \sum_{j=i+1}^{k+1} \sum_{v \in \mathit{Drains}(\vec{H}_j)} \text{Correction-Dist}_{\vec{H}_j}(v). \qquad (4)$$

$\square$

B.2 Proof of Lemma 5.7

For the proof of the Lemma, we need the following notations. Let $L_{\vec{G}}(\epsilon) = \bigcup_{v \in \mathit{Drains}(G) \setminus \mathit{Heavy}} \tilde{B}(v, \mathit{rad}(v))$, $\overline{L_{\vec{G}}}(\epsilon) = G \setminus L_{\vec{G}}(\epsilon)$. For every two G-orientations $\vec{G}_1, \vec{G}_2$, denote by $\overline{\mathit{dif}}(\vec{G}_1, \vec{G}_2)$ the number of edges in $\overline{L_{\vec{G}}(\epsilon)}$ such that their direction in $\vec{G}_1$ is different than their direction in $\vec{G}_2$. That is, $\overline{\mathit{dif}}(\vec{G}_1, \vec{G}_2) = \{\{u,v\} \in E\left(\overline{L_{\vec{G}}(\epsilon)}\right) : (u,v) \in E(\vec{G}_1), (v,u) \in E(\vec{G}_2)\}$.

To prove the Lemma it is enough to prove that for every $\vec{G} \in$ 2DSF such that $\vec{G}$ is ϵ-far from Drain-Free, there exists a G-orientation $\vec{G}' \in$ Drain-Free, such that $\overline{\mathit{dif}}(\vec{G}, \vec{G}') \leq \frac{\epsilon |E(G)|}{6}$. Since $\vec{G}$ is ϵ-far from Drain-Free, the claim follows. Let $\vec{G}$ be as described above.

We construct the orientation $\vec{G}'$ in two stages as follows.

Stage 1: we first construct a G-orientation $\vec{G}_1$ as follows: Set $\vec{G}_1 = \vec{G}$ and $S = \mathit{Drains}(\vec{G}_1)$. For every $v \in S$, according to some ordering of S, do the following. If the path Correction-Path$_{\vec{G}_1}(v)$ is fully contained in $\tilde{B}_G(v, \mathit{rad}(v))$, flip all the edges in Correction-Path$_{\vec{G}_1}(v)$. Otherwise, if there is a heavy vertex $x \in \tilde{B}_G(v, \mathit{rad}(v))$, direct towards x all edges on a shortest path connecting v and x. It is important to note that for each $v \in S$ flipping the edges along Correction-Path$_{\vec{G}}(v)$ is according to the current state of $\vec{G}_1$, that is the state after the previous flipping. Since $B(v, \mathit{rad}(v))$ are vertex disjoint a drain can be created only if it is a heavy vertex, thus $\mathit{Drains}(\vec{G}_1) \setminus \mathit{Heavy} \subseteq \mathit{Drains}(\vec{G}) \setminus \mathit{Heavy}$.

Observe that for the resulting orientation $\vec{G_1}$ and for every $v \in \mathit{Drains}(\vec{G_1}) \setminus \mathit{Heavy}$ we have Correction-Dist$_{\vec{G}}(v) > \mathit{rad}(v)$ and hence $\tilde{B}_{\vec{G_1}}(v, \mathit{rad}(v))$ is fully contained in $\tilde{B}_G(v, \text{Correction-Path}_{\vec{G_1}}(v) - 1)$. Therefore, by Claim B.3, for every vertex $v \in \mathit{Drains}(\vec{G_1}) \setminus \mathit{Heavy}$ we have that $\tilde{B}_G(v, \mathit{rad}(v))$ is a tree and all the edges are directed towards v. In addition, by Claim B.4 for every pair of different vertices $u, v \in \mathit{Drains}(\vec{G_1}) \setminus \mathit{Heavy}$ we have that $\tilde{B}_G(u, \mathit{rad}(u))$ and $\tilde{B}_G(v, \mathit{rad}(v))$ are vertex disjoint. Most importantly, by the definition of $\mathit{rad}(v)$ and the definition of the correction path as a shortest path to some target vertex, for every $v \in \mathit{Drains}(\vec{G_1}) \setminus \mathit{Heavy}$ there are at least $36/\epsilon$ vertices at distance $\mathit{rad}(v)$ from v.

Stage 2: Define $\vec{H}$ to be the graph obtained from $\vec{G_1}$ in the following manner: For every $v \in \mathit{Drains}(\vec{G_1}) \setminus \mathit{Heavy}$, according to an arbitrary ordering, contract $\tilde{B}_G(v, \mathit{rad}(v) - 1)$. Note that the subgraphs contracted are pairwise vertex disjoint.

We next show that $\mathit{Drains}(\vec{H}) \subseteq \mathit{Heavy}_H$. Observe that $\mathit{Drains}(\vec{H})$ contains vertices that are in $\mathit{Drains}(\vec{H}) \cap \mathit{Heavy}_G$ and vertices and that are a result of the contractions. Since $\tilde{B}_{\vec{G_1}}(v, \mathit{rad}(v))$ is a directed tree whose edges are directed towards v and has at least $36/\epsilon$ ingoing edges, the contraction transforms $\tilde{B}_{\vec{G_1}}(v, \mathit{rad}(v) - 1)$ into a heavy drain.

By lemma 5.8, there exists $\vec{H}' \in \text{Drain-Free}_H$ such that $\mathit{dist}(\vec{H}, \vec{H}') \leq \frac{4|E(G)|}{\Delta_{Dr}(\vec{H}}) \leq \frac{\epsilon|E(G)|}{9}$.

The G-orientation $\vec{G_2}$ is obtained from $\vec{G_1}$ by directing all edges that correspond to edges in $\vec{H}'$ in the same way that the corresponding edges in $\vec{H}'$ are directed. Based on what was argued before, $\mathit{dist}(\vec{G_1}, \vec{G_2}) \leq \epsilon|E(G)|/9$, and hence $\overline{\mathit{dif}}(\vec{G_1}, \vec{G_2}) \leq \epsilon|E(G)|/9$.

Note that there are no heavy drains in $\vec{G_2}$ since all the heavy vertices in $V(G)$ are also in vertices in $V(H)$ and moreover, all the edges connected to them are directed the same way in $\vec{G_2}$ and in $\vec{H}'$. Note also, that all drains in $\vec{G_2}$ are also drains in $\vec{G_1}$; namely, those are drains $v \in \mathit{Drains}(\vec{G_1})$ that where not corrected in step 1 and now for every such $v \in \vec{G_2}$ there is a target vertex in $\tilde{B}_{\vec{G_2}}(v, \mathit{rad}(v))$. Thus, since the $\tilde{B}_G(v, \mathit{rad}(v))$ are vertex disjoint for different $v \in \mathit{Drains}(\vec{G_2})$, by flipping the edges in Correction-Path$_{\vec{G_2}}(v)$ for every remaining drain $v \in \vec{G_2}$, we get a drain-free orientation. $\square$

C Test for Drain-Source-Free

We present now the test for general inputs (relaxing the assumption of being $2DSF$). We need the following definition.

Definition C.1. *For every edge $e = \{u, v\} \in E(G)$ define P_e to be the maximal length path using e such that all the internal nodes of P_e have degree 2. That is, P_e is a path between two (not necessarily different) vertices x and y such that the following holds: (a) P_e uses the edge e, (b) $\mathit{deg}_G(x) \neq 2$, $\mathit{deg}_G(y) \neq 2$, and (c) $\mathit{deg}_G(z) = 2$ for every vertex $z \neq x, y$ in P_e. For a G-orientation $\vec{G}$ denote by $\vec{P_e}$ the subgraph of $\vec{G}$ that corresponds to P_e. We say that $\vec{P_e}$ is consistent if it is a directed path.*

Note that if a G-orientation has an edge e such that $\vec{P_e}$ is not consistent then it has a drain or a source of degree 2.

The new algorithm is obtained from Algorithm 5.1 by replacing each query of an edge e by random sampling of the path P_e. A G-orientation is rejected if either a source or a drain were detected, or if one of the paths P_e sampled is not consistent. The algorithm is now defined as follows:

Algorithm C.1.
Input: $\epsilon, \vec{G}$.
Repeat $48/\epsilon$ *times:*

1. *Randomly choose an edge* $e = \{u, v\} \in E(G)$ *such that* $v \notin Heavy$.

2. *Select* $48/\epsilon^2$ *edges uniformly and independently from* P_e.

3. *If* P_e *is not consistent then return REJECT.*

4. *For every relevant vertex* z *such that* $dist(v, z) \leq rad(z)$, *do the following.*

   - *For every edge* $e' = \{z, x\} \in E(G)$ *select* $48/\epsilon^2$ *edges uniformly and independently from* $P_{e'}$ *and query their direction as well as the direction of* e'.
   - *If for one of the edges* e', $P_{e'}$ *is not consistent then return REJECT.*
   - *If a drain or a source is found then return REJECT.*

5. *If there was no rejection in previous steps then return PASS.*

As was previously noted, this algorithm is similar to algorithm 5.1 with the following two differences. Each query in the algorithm is replaced by sampling some random edges in the related path $P(e)$ and the new uses the notion of a consistent path.

Lemma C.2. *Algorithm C.1 is a* $(\epsilon, (1/\epsilon)^{O((\frac{1}{\epsilon})^2)})$*-test for Drain-Source-Free.*

Proof. Clearly, the number of queries used by algorithm C.1 is $3 \cdot 48/\epsilon^2$ times the number of queries used by algorithm 5.1 and hence it is as claimed. Since the algorithm only rejects if there is a drain/source in the orientation the algorithm has a one-sided error. Thus, it remains to show that algorithm C.1 rejects every G-orientation that is ϵ-far from Drain-Source-Free with probability of at least $\frac{2}{3}$.

Let $\vec{G}$ be ϵ-far from Drain-Source-Free and let $\vec{G}' \in$ 2DSF be such that $dist(\vec{G}, \vec{G}') = dist(\vec{G}, \text{2DSF})$. We distinguish between two cases.

1. $\vec{G}$ is $\epsilon/6$-far from 2DSF. Define $Path(G)$ to be the set of all paths P_e that contain at least one vertex z such that $deg_G(z) = 2$. Observe that $dist(\vec{G}, \text{2DSF}_G) = \sum_{P \in Path(G)} dist(\vec{P}, \text{2DSF}_P)$. By averaging considerations, there exists a set $\mathcal{T} \subseteq Path(G)$ such that $\sum_{P \in \mathcal{T}} |E(\vec{P})| \geq \epsilon \cdot \epsilon |E(G)|/12$ and for every $P \in \mathcal{T}$ we have $dist(\vec{P}, \text{2DSF}_P) \geq \epsilon \cdot |E(\vec{P})|/12$. The probability that $P_e \in \mathcal{T}$ for an edge e selected in step 1 of algorithm C.1 is at least $\epsilon/12$. The probability that such an e is selected in one of the iterations is at least $(1 - \epsilon/12)^{48/\epsilon} > 9/10$. For such P_e the probability that 2 edges of opposite direction are not selected in step 2 of algorithm C.1 is at most $(1 - \epsilon/12)^{48/\epsilon^2} + (\epsilon/12)^{48/\epsilon^2} < 1/16$. Hence the probability that algorithm C.1 rejects is at least $(1 - 1/16)9/10 > 2/3$.

2. $\vec{G}$ is $\epsilon/6$-close to 2DSF. By the triangle inequality, $\vec{G}'$ is $2\epsilon/3$-far from Drain-Source-Free. Consider a relevant vertex $z \in Drains(\vec{G}')$ that is not a drain or a source in $\vec{G}$. Since z is not a drain or a source in $\vec{G}$ then for one of the edges $e = (z, x)$ the direction of the edge e in $\vec{G}$ is not consistent with the direction of $\vec{P}_e$ in $\vec{G}'$. By the definition of $\vec{G}'$, at least half of the edges in P_e are directed as in $\vec{G}'$ and hence are not consistent with the direction of e. Thus, sampling each of these edges we either find a drain or a source or we find an inconsistent

path. Given such a relevant vertex z, the probability that the algorithm failed to find a drain or a source or a not consistent path is bounded by $\frac{1}{2}^{\frac{48}{\epsilon^2}}$.

By the correctness proof of Algorithm 5.1, the probability that for some relevant vertex z that is found in step 4 we have that $z \in \mathit{Drains}(\vec{G}')$ is at least 9/10. By the above discussion, the probability that the algorithm finds a drain or a source or a not consistent path for such a vertex v is also at least 9/10. Hence, the algorithm rejects $\vec{G}$ with probability of at least 2/3.

$\square$

D Proof of Theorem 7.4

Proof sketch: By Theorem 7.3 it is enough to show that the $\sqrt{n} \times \sqrt{n}$ grid is strongly-Steiner-connected.

Indeed, let G be the 2-dimensional grid and let $S \subseteq G$ with $|S| \leq \delta^2 n$. Divide G into vertex disjoint subgrids each of size $\frac{1}{\delta^2} \times \frac{1}{\delta^2}$ by partitioning the rows and the columns of G into consecutive blocks of size $\frac{1}{\delta^2}$. Let P be a path going through all the centers of these blocks, whose length is at most $\delta^2 n$ (it is easy to see that such a path exists). We can connect each point in S by a path of length at most $2/\delta$ to the closest center thus creating a connected subgraph that spans S. Obviously, as $|S| \leq \delta^2 n$ the total number of edges in this subgraph is bounded by $3\delta n$. $\square$

E Proof of Observation 6.2

Let $V' = \{v | v \text{ appears in a copy of } H \text{ in } \vec{G}\}$. We will show that $dist(\vec{G}, H\text{-free}) \leq |V'| \cdot \Delta$. Indeed, pick an arbitrary vertex $v \in V'$ and direct all its incident edges to be outgoing. This obviously makes it impossible for v to belong to any H-copy in the resulting orientation as H is source-free, and in addition, it creates no new copy of H. This can now be iterated on the other vertices of V'. At each step, at most Δ-edges are flipped, while finally the resulting graph is H-free. $\square$

F Proof of Theorem 6.3

In this section, we prove that the second item indeed holds for hard to test coloring. By this, we complete the proof of Theorem 6.3.

Assume that $c : V(G) \longrightarrow \{0, 1, 2\}$ is ϵ-far from being a proper 3-coloring and contains $o(n)$ monochromatic edges. Let C_0, C_1, C_2 be the corresponding color classes. Let $\vec{G}_c$ be the corresponding orientation and assume that $dist(\vec{G}_c, P_3 - free) = \beta n$. This means that we can redirect at most βn edges so to get a P_3-free orientation $\vec{G}^*$. We now define a proper 3-coloring of G based on $\vec{G}^*$. Let B be all the vertices that are adjacent to an edge that is oriented differently in $\vec{G}_c$ and $\vec{G}^*$. Note that $|B| \leq 2\beta n$.

Let $U_0 = \{v \mid v \; is \; a \; start \; vertex \; in \; a \; 2-path \; with \; respect \; to \; \vec{G}^*\}$. Let $U_2 = \{v \mid v \; is \; a \; end \; vertex \; in \; a \; 2-path \; with \; respect \; to \; \vec{G}^*\}$. Let $U_1 = \{v \mid v \; is \; a \; middle \; vertex \; in \; a \; 2- path \; with \; respect \; to \; \vec{G}^*\}$. For other vertices: for $u \in B$ we put u in U_0 if u is a source w.r.t $\vec{G}^*$ and in U_2 if it is a drain w.r.t. $\vec{G}^*$. For $u \notin B$ we put u in U_0 if it is a source and either $u \in C_0$ or $u \in C_1$ but has a neighbor in C_1, if u is a source but has no neighbor in C_1 u goes to U_1. Analogously, we put u that is a drain in U_2 if it is in B or it is in C_2 or it is in C_1 and has a neighbor in C_1. Otherwise, we put it in C_1.

We claim that the partition obtained is consistent with $\vec{G}^*$ and thus defines a proper coloring c^*. To see that the only thing to be checked is that no two vertices in $U_1 - B$ are neighbors. Indeed this immediately follows, as otherwise, either they are both middle points of a 2-path, in which case there is a 3-path in $\vec{G}^*$ in contradiction to its definition. Or one is source and one is a drain belonging to C_1 with no neighbors in C_1 which contradicts the fact that they are neighbors.

It remains to show that the coloring obtained c^* is 'close' to c. Indeed, let us analyze the number of recolored 0's. That is, the number of vertices in $U_0 - C_0$. Each such vertex u is either in B of which there are at most $2\beta n$ vertices, or it is a source in $\vec{G}$, it belonged to C_1 and has a neighbor in C_1 - however, by assumption on c there are $o(n)$ such vertices. A similar bound applies to $U_2 - C_2$. Finally, $u \in U_1 - C_1$ if it either belongs to B or it belongs to $C_0 \cup C_2$ and is, or became a middle point in a 2-path. Note that there are at most $2\beta n$ vertices for the first case. If, $u \in C_0$ and is a mid point of a 2-path $v - u - w$ with respect to c, it means that $v \in C_0$ and thus there are only $o(n)$ such vertices. If $u \in C_0$ and became a mid point of a 2-path $v - u - w$ it means that either $v \in B$ or $w \in B$ but as G has degree bounded by Δ there are at most $4 \cdot \Delta\beta n$ such vertices. A similar calculation is valid for the other final case of $u \in C_2$. Hence, altogether $dist(c, c^*) \leq (8 \cdot \Delta + 6)\beta n + o(n)$. This implies that $\beta > \epsilon/(12\Delta)$ which completes the proof. $\square$

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'